

Chapter Title: Against perfectionism

Book Title: Hedonistic Utilitarianism

Book Author(s): Torbjörn Tännsjö

Published by: Edinburgh University Press. (1998)

Stable URL: <https://www.jstor.org/stable/10.3366/j.ctvxcrx5b.10>

---

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

*Edinburgh University Press* is collaborating with JSTOR to digitize, preserve and extend access to *Hedonistic Utilitarianism*

### *Against perfectionism*

I have argued that both hedonism and perfectionism are more plausible theories of value than preferentialism. I have also indicated that I find hedonism more plausible than perfectionism. I have not argued this point, however. This will be the object of the present chapter. My argument will be directed at what I find to be the most plausible version or aspect of perfectionism, the idea that what matters to a person is that his or her life be genuinely his or hers, that he or she be autonomous. This seems, moreover, to be a common prerequisite of various other different ideas of perfection, such as the idea that achievements are of value to us as such, or the idea that knowledge is of value to us as such, or the idea that deep personal relations or a contact with realities are of value to us as such. What these ideas share is the common assumption that, if an achievement of mine, or a piece of knowledge of mine, or a personal relation of mine is to be of value, no matter what exact content it may have, it must be my own.

The claim that autonomy has intrinsic value will be the immediate target of the argument of this chapter. The thrust of this chapter, then, is that personal autonomy is of no value as such. Personal autonomy is not a basic moral category (which does not preclude, of course, that it may often be a means to something which has value, i.e., pleasurable experiences). If this argument is sound, and if my observation that autonomy is a prerequisite of all interesting forms of perfection is correct, then, indirectly, perfectionism as such is being defeated as well.

#### Two contexts for autonomy

Concepts of autonomy are invoked mainly in two kinds of moral context. On the one hand, it is sometimes claimed autonomy is something which is required from us. This is the view put forward by Kant. We ought to be autonomous.<sup>1</sup> On the other hand, autonomy (or individuality, to use J. S. Mill's word)<sup>2</sup> is something we could invoke to protect ourselves against certain kinds of unjust demand: our autonomy or individuality is something which others ought to respect. We have a right to autonomy, or, our autonomy (individuality) is something which is of value to us and, therefore, of value in itself.

In this chapter I will discuss the concept of autonomy which is adequate in the latter context. Is there a tenable ideal of personal autonomy? My main thesis is that there is no such tenable ideal. It has no importance as such whether an action we perform restrains the personal autonomy of someone or not. Hence, it would not be a good argument after all against, say, manipulation of behaviour if it could be proved that it necessarily involves restrictions of personal autonomy.

### The concept of autonomy

What does it mean to say of a person that she or he is 'autonomous'? In the present context the most plausible answer would be that she or he is a person who does what she or he chooses or decides to do, and who makes the choices and decisions she or he does because she or he wills it. Although this concept of autonomy is of course far from clear, I think it could still easily be made sufficiently precise for the purpose of this chapter.

In the first place, it is clear that the concept could be taken either in a straightforwardly empirical way or in some theoretical, ideal way. Are the actual will and decisions of a person what matter to her or his autonomy, or the ideal ones which she or he might have had or manifested if she or he had been, in some respect, different?

The use of ideal concepts of autonomy is very common indeed among those who follow Kant and argue that we ought to be autonomous. John Rawls, for example, writes: 'Thus acting autonomously is acting from principles that we would consent to as free and equal rational beings, and that we are to understand in this way.'<sup>3</sup> Philosophers who want to follow J. S. Mill tend rather to use the empirical conception of autonomy even when they depart from his terminology and use the word 'autonomy' (Mill spoke instead of individuality or self-development). Jonathan Glover, for example, writes:

You must have the desire whose satisfaction is in question. I override your autonomy only where I take a decision on your behalf which goes against what you actually do want, not where the decision goes against what you would want if you were more knowledgeable or more intelligent.<sup>4</sup>

As was pointed out by J. S. Mill, both conceptions of autonomy have deep historical roots and are sometimes difficult to separate from each other.<sup>5</sup> I will examine both concepts, taking as my point of departure an empirical conception and postponing until later the scrutiny of various ideal ones.

It is important to notice that the actions of autonomous persons are in accordance with both their decisions and their wants. Their actions track their decisions and their decisions track their wants.

It is true that the psychology taken for granted in my definition is crude. In the

final analysis, no matter how we should like to account for such things as a person's will, his or her choices or decisions and his or her actions, these are, I think, roughly the distinctions that we should like to be able to make. I am taking some unspecified causal theory of action for granted and using the word 'will' in a slightly technical sense. My will in a decision context may be at odds with several of my desires. We give greater weight to those desires that are stronger, although one strong desire may be outweighed by several other desires. What I want to do in the situation, my will in the situation, is what I believe would best fulfil all my desires. A person who 'wants', in this sense, to do one thing but who does something else is not autonomous.

In what follows I will defend the view that such weakness of the will is possible. The causal relation between what a person wants and what he or she decides to do, and between what he or she decides to do and what he or she does, must be of the 'right' or 'standard' kind, if the person's action is to be said to be autonomous. However, I cannot go into the problem here of how this is to be understood more exactly.

To begin with, let us say that persons are autonomous if they do what they decide or choose to do and do this because they actually want to do it. Their autonomy is restrained if they are hindered by some person or authority from doing what they decide or choose to do because they actually want to do it. This means that while hypnosis may present us with good examples of violations of our autonomy, it is more difficult to see that manipulation of our situation, restraints on our alternatives and so forth may equally threaten it. Autonomy is one thing, freedom another.<sup>6</sup> Even if only two alternatives are left open to an agent, to commit treason, for instance, or undergo torture, he or she may choose autonomously between these. It is also far from clear how blind obedience may threaten our autonomy. If I choose to obey and want to choose to obey, then I am also autonomous when acting in accord with my obedience.

### The ideal of personal autonomy

Persons are autonomous if they do what they decide to do and decide to do what they decide to do because they want to do it: i.e., if their actions track their decisions, and their decisions track their wants. What, then, is the ideal of autonomy? Is it simply that people should be allowed to be autonomous? No; we ought to distinguish various aspects of autonomy. If we have a duty to be autonomous, this may be a duty always to be autonomous, but the view that we have a right to have our autonomy respected is most plausible if it is confined to what could be called personal autonomy. You could be autonomous, then, with respect to your choice of occupation, sexual partners, newspapers and so forth. The respect dearest to the adherents of the ideal of autonomy is probably such personal choices. This is also what is at the heart of Robert Nozick's idea of self-

ownership in *Anarchy, State, and Utopia*. I will take the ideal of (personal) autonomy primarily to include the idea that (at least) persons should be allowed to do what they want to do, and decide to do, with themselves.

Nevertheless, is it not obviously unreasonable to maintain that people should be allowed to do what they want to do and decide to do, even with themselves? For the self-perfection (or self-destruction, for that matter) of one person may take place at the expense of some other person. It is not reasonable to hold that one person should be allowed to use the scarce resources of a society to develop her or his talent as a violin player when her or his compatriots are actually starving to death, if this would-be violinist might be an efficient farmer, as well. Yes, this is reasonable, some would say (such as Nozick), while others would say no. This is controversial, then. So in order to make the ideal as plausible as possible, I will take it to imply only that no one should be manipulated into doing what she or he does not want to do, or does not decide to do, with herself or himself, when this manipulation is in her or his own interest alone and no one else's.

The ideal of personal autonomy vaguely set forth here could be understood in at least three different ways. First of all, it could be understood as an absolute norm, one which should never have to give way to any other consideration, and which dictates that personal autonomy must never be restrained. It could also be interpreted to mean that the value of personal autonomy is infinite, or that the value of personal autonomy should be observed fully, before any other values are considered at all. To use a phrase from John Rawls: the value of personal autonomy is lexically prior to all other principles.<sup>7</sup>

Secondly, it could be understood to mean that it is a matter of negative intrinsic value whenever the autonomy of a person is restrained, even if in some cases it might be right to so restrain her or his autonomy (the consideration of personal autonomy is outweighed by some other consideration, say, for the happiness of this person in the long run). This is how James Griffin conceives of the ideal. According to Griffin, even 'if I constantly made a mess of my life, even if you could do better if you took charge, I would not let you do it. Autonomy has a value of its own'.<sup>8</sup> And yet, for all that, autonomy, 'should not be realized . . . when it sets up great anxieties in a particular person'.<sup>9</sup>

Thirdly, the proscription against interference with personal autonomy can be justified, not as a matter of abstract rights, or because of the intrinsic value of personal autonomy, but rather because of the overall gain in net utility obtained by universal recognition of such an absolute value (rule utilitarianism).

### Is there an absolute right to personal autonomy?

Admittedly, there are many cases where the personal autonomy of someone is violated and where this is wrong. The person who gets his or her autonomy

tampered with may be hurt. Or, being of a Kantian bent, he or she may feel that, when we succeed in manipulating him or her, he or she fails to live up to the standards he or she has set his or her own life. However, these examples do not show that violations of personal autonomy are wrong as such. And my thesis is that no examples that do show this can be found. We have no absolute right to personal autonomy. My present claim is that, even if personal autonomy may be of some value as such, it is at least not of infinite value.

In order to make plausible the claim that we have no absolute right to personal autonomy I need a clear example to which I can refer. What is required is a case where a person is not allowed to do what he or she wants and decides to do with himself or herself only because what he or she wants and decides to do with himself or herself is not considered to be good for him or her, and where it is nevertheless fairly obvious that it would be right not to allow him or her to do what he or she wants and decides to do with himself or herself. How could such an example be constructed? To be convincing it should be a case where the person in question has a very strong and permanent desire to do what he or she is not allowed to do. And the action he or she wants to perform must concern something he or she considers very important to his or her ego. It would not be sufficient for him or her not to attain what he or she wants and decides. He or she must be actively manipulated to do what he or she wants not to do or has not decided to do. I think the following example will serve the purpose.

Consider a man who is going to die of lung cancer. He gets little relief from the morphine he is given by his doctor, who has cut down his dose in order not to make his breathing more difficult. The doctor asks him whether he would like to have a lethal dose of the morphine. The patient refuses. To be sure, the rest of his life, consisting as it does of nothing but severe physical displeasure and discomfort until he is suffocated by his disease, will not be worth living. Although he realises this, and hopes that death will follow soon, for moral reasons he believes that it is absolutely forbidden to take a human life, or to assist when a human life is taken. That is the belief he has come to accept after having seriously considered the problem of euthanasia. To press matters, we may even assume that this person has made fidelity to religious duties and rituals the whole pattern and meaning of his life. And the causes of this religious attitude, as taken up by the patient, may be fully known to him. It is not possible, we assume, to undermine it through any kind of 'cognitive psychotherapy' of the kind mentioned in the previous chapter. Apparently, the doctor accepts his refusal. However, the doctor secretly gives the patient an overdose, thus killing him.

Let us suppose that, for one reason or another, there were just two rough options open to the doctor. She could, as she did, manipulate the patient into accepting a lethal dose, or she could give him the usual dose, but she could not, for example, force the lethal dose on the patient (which few would consider the best option in the situation, anyway). It is my considered opinion that the doctor

did the right thing (let us also take it for granted that there are no bad side-effects of her action). This also seems to be the conclusion we must draw if hedonistic utilitarianism is brought to bear on this case. How could I convince those who have their doubts about this?

Perhaps most people have their doubts about it. Perhaps they would say that, in examples like the present one, the patient should not be given a lethal dose. I concede that this may be so. But I think our 'intuitions' about cases like the present one are much too biased and prejudiced to be taken at face value. We are not capable of keeping our minds clear and our heads cool when we consider cases such as this. In particular, we tend to forget that what is here being discussed is a mere thought-experiment, where there is no doubt (as there must always be in real life) that there are no bad side-effects of an act of involuntary euthanasia (not to speak of doubts concerning the value of the patient's remaining life). We should approach our intuitions more indirectly. It would probably be a good idea to take as our point of departure our intuitions about more abstract problems, such as whether we ought to avoid unnecessary displeasure or whether persons ought to be allowed to do as they please with themselves. Most people agree that we ought to avoid unnecessary pain. It is controversial, however, whether people have a right to do as they please with themselves. Since we are facing a conflict here, some intuitions will have to yield. And a certain onus rests on the person who wants to argue that, sometimes, we need not avoid avoidable pain.

If the doctor has the patient killed she assists in an act of involuntary euthanasia. By manipulating the patient into accepting the lethal dose, moreover, she restrains the personal autonomy of the patient. The patient accepts the injection only because he wrongly believes he will get his ordinary dose, but instead gets a lethal dose. Now, why should it be wrong of the doctor to do this, if it is in the interest of the patient to have the lethal dose and not in conflict with the interest of anyone else?

Peter Singer has argued, against involuntary euthanasia, that it presupposes that 'one can judge when a person's life is so bad as to be not worth living, better than the person can judge herself'.<sup>10</sup> Singer believes that it:

is not clear that we are ever justified in having much confidence in our judgements about whether the life of another person is, to that person, worth living. That the other person wishes to go on living is good evidence that her life is worth living. What better evidence could there be?<sup>11</sup>

In relation to my example it is clear that Singer's objection fails. That the patient in my example (my thought-experiment) wants to go on living, and decides to turn the offer of euthanasia down, is not evidence at all, and certainly not good evidence, that his life is worth living. He wants to continue to live

solely for moral (religious) reasons. And our judgement that his life is not worth living does not presuppose that we know this better than he does. As a matter of fact, he shares our opinion on this point. He admits that what remains of his life is not worth experiencing. He hopes that he will soon die. He, however, does not want to put an end to his own life. He feels that, if he did, it would mean that his entire life would lose its point. However, if the doctor succeeds in manipulating the patient, no harm is done to the patient (even by his own lights). To be sure, the patient kills himself, or at least accepts being killed, but the patient cannot reasonably be blamed for what he does. If anyone is to receive blame, it is the doctor. So the life of the patient retains whatever point it would have had, by his own lights, had he not been manipulated into accepting a lethal dose. Furthermore, the patient is spared unnecessary pain.

The previous example rebuts only one common argument against involuntary euthanasia. However, there also exists one consideration which could be taken into account that supports strongly the view that there is no tenable absolute ideal of personal autonomy. This consideration is the fact, and I think it is a fact, that we tend to disregard the ideal of personal autonomy when two requirements are satisfied. These requirements are, first of all, that by restraining someone's personal autonomy we can spare the person extreme pain, and, secondly, that the person concerned is someone who is dear to us, someone we really care about. There is a telling example of this in Jonathan Glover's book, *Causing Death and Saving Lives*. Although Glover argues in defence of an ideal of personal autonomy, he admits that there may be exceptional cases:

It does not seem plausible to say that there is no *conceivable* amount of future misery that would justify killing someone against his will. If I had been a Jew in Nazi Germany, I would have considered very seriously killing myself and my family, if there was no other escape from the death camps. And, if someone in that position felt that his family did not understand what the future would feel like and so killed them against their wishes, I at least am not sure that this decision would be wrong.<sup>12</sup>

I believe that most of us share Glover's feelings. However, feelings are one thing, a considered moral opinion something else. Therefore, how are we to interpret Glover's example? One interpretation would be that, in situations such as the one described, we fail to make a considered moral assessment; our passionate feelings lead us astray. No moral conclusion can validly be drawn from the example. This position is difficult to sustain consistently, however. Why should we assume that our feelings have led us astray? To be sure, the example is construed so as to guarantee that we are maximally involved. And a strong emotional involvement is sometimes a sign of a bad judgement. However, our emotional involvement does not *per se* distort our judgement. An emotional involvement may prepare the ground for rash assessments



when it carries a bias. In the example under discussion, however, we, who conduct the thought-experiment, should not be involved in any way that is essentially partial. The main effect of our involvement is different. Because we are so emotionally involved, we fail to detach ourselves from the problem. Had strangers been involved we might have felt that, even if we allow them to choose options that are extremely painful for them, we are still not really responsible for their suffering. In relation to strangers, because of our lack of empathy, we succeed in sticking to the abstract ideal of personal autonomy and we allow them to inflict displeasure on themselves. This attitude is certainly much easier to take up if strangers are involved. But are we any less responsible for their displeasure than we would be if those who suffered were near and dear to us? I would say that we are not. We feel this if we are deeply involved in the case. And I would say that our involved (but impartial) judgement should be given more weight than our detached one. It is a fact, or so I believe, that for almost each and any one of us, there is some amount of suffering such that, if we could stop those who are near and dear to us from voluntarily inflicting it on themselves, we would do so. This fact could reasonably be taken to indicate that there are situations where it would be positively wrong to allow even strangers to choose such options.

If, however, it is true that sometimes we are allowed to kill people against their will, there must be all sorts of situation where, in order to save people from severe pain, it is all right to restrain their personal autonomy in less serious respects. If this conclusion is sound, it means that there is no absolute right to personal autonomy. The value of personal autonomy, if it is of value in itself, is not infinite.

### Are autonomous actions of value in themselves?

There is no absolute right to personal autonomy, or so I have argued at any rate. The possibility may, nevertheless, remain that there is some value in having one's personal autonomy respected, even if this value must sometimes yield to other values.

I am not quite sure how the view that personal autonomy possesses value in itself can be presented as plausibly as possible. The point cannot really be that there is some positive intrinsic value in each autonomous action. It would be strange if we could make the world better by performing as many actions as possible. Instead, there may be some negative value involved in each case where we are manipulated or otherwise made to perform an action which is not autonomous. On the other hand, the argument may be that it is good in itself to be an autonomous person, no matter how often the autonomy is exercised, i.e., a person who is capable of performing autonomous actions. According to the latter argument, if our autonomy is taken away from us (or if it is voluntarily

given up by us or just, say, because of bad luck, lost), then something of value in our lives is lost. I should like to examine both possibilities.

Consider first a person who has retired from his occupation as a carpenter. He leads a quiet life in a small town which he never leaves. He is enjoying the last few years of his life doing some gardening and reading some philosophy. As far as he is concerned, his life lacks only one thing. While he would like to see his two children twice a year, each of them only visits him once a year. There is the possibility that he could visit them sometimes. They would make him feel welcome and he would enjoy staying with them for a while. He dares not make the journey, however, for fear of getting killed in some kind of accident. As a matter of fact, he never travels by car, train or aeroplane. His problem is not that he has an unrealistic view of the probability of getting killed. He knows that the chance of getting killed if he travelled once a year is in fact very small. He has none the less come to entertain the idea that nothing could be so good that it could provide a reasonable motive for him to increase the risk of getting killed in an accident by however small an amount. After all, the improvement in his life would be marginal if he saw his children twice a year instead of just once. He feels, therefore, no inclination to take the risk of going to see them.

One day a friend of this man, a retired doctor, comes to see him and offers him a new medicine. 'Take it', she says. 'It will make you react as people normally do to the risk of getting killed in an accident. The medicine will not change your view of the probabilities, but it will make you react to them in a more relaxed way. The medicine is in itself quite harmless.'

The man's answer is that he does not want to take the medicine. 'I dare not take it', he says. 'It would obviously make me a person who is prepared to gamble to some extent with life. The expected value of my life would be slightly better if I took it. I could then see my children twice a year, which is better than once a year, and it is highly improbable that I should get killed in an accident. But if I take the medicine I will go and see my children and this means that the risk that I get killed in an accident will increase somewhat. And I am not willing to do anything which increases in any way the risk of getting killed.'

The doctor, knowing how stubborn her friend is, says nothing more, but makes a birthday cake for her friend, puts the medicine into the cake and gives it to him. It does not take long before the man realises what has happened. He eats the cake and all of a sudden views the prospect of going to visit his children without horror. Then he goes to his friend, the doctor, and thanks her. 'But does it not disturb you', the doctor asks, 'that I have changed your personality against your will?' 'Not in the least', the man answers. 'I, speaking as the person you have turned me into, have no objections. Now I am capable of living a life which is slightly better than the one I, the person that used to be me, used to live.'

In the face of this example it is possible for me to assert without hesitation that the doctor did the right thing. It does not matter how little the increase in value

of the life of the man was, as long as there was some improvement (and no bad side-effects). This being so, the doctor did the right thing. This conclusion, if correct, would prove that there is no value in itself in having one's will with regard to one's life respected. Is it, however, correct? Here some of the possible objections to my conclusion should be examined.

It could perhaps be suggested that there is inevitably some negative value in having one's wants frustrated. It does not feel good not to get what one wants. This argument is based on an equivocation, however. The word 'frustration' is ambiguous. It can be used either to refer to an experience (which is, more often than not, unpleasant) or to refer to an abstract relation (which cannot be felt or experienced at all). In my example above the 'frustration', if there is any, is of the abstract kind. It consists of a lack of correspondence between what is at one time wanted and what is later the case. Nothing in the example indicates, however, that the realisation of this discrepancy is followed by any unpleasant feelings on the part of the person who took the medicine. On the contrary, he says after he has taken the medicine that he is glad his (former) will was not respected. And the 'former person', who would have complained if he had 'survived' the taking of the medicine, does not exist any longer, so he cannot complain and he cannot suffer from any loss. Hence, no frustration is felt in the example.

If, however, there is nothing wrong in changing a person's character, as long as the result is that she or he will not object afterwards to the change, does not this mean that it would be unobjectionable as such to exchange one person for another?

Perhaps it does. The distinction is not quite clear between a change within a person, on the one hand, and an exchange of one person for another, on the other. The concept of a person is not absolutely definite. Except in some fanciful thought-experiments, we seldom have difficulties in identifying actual persons. Each person occupies a certain place in space-time. But it is difficult to tell what it is about such an entity in space-time that makes it a person and that makes it one person. Some would stress physical continuity, others would concentrate on mental traits, such as connectedness by memory, 'projects' and the like. Most would opt for a combination. Physical continuity and mental connectedness, however, come in degrees. It is not easy, therefore, to say when one person has been changed and when one person has been exchanged for another. So, probably, if there is nothing which makes it wrong in principle to change a person against her or his will, there is probably nothing which makes it wrong in principle to exchange one person for another. If we want to defend the former claim we should be prepared to defend the latter.

Bernard Williams argues that 'it is absurd' to demand of a person that he or she give up his or her projects which he or she is most deeply and extensively identified with. Why should this be so? The reason is, according to Williams,

that it is 'to alienate him in a real sense from his actions and the source of his actions in his own convictions'.<sup>13</sup>

Let us for the sake of argument concede that to change a person's personality in a way which threatens his or her personal autonomy is to alienate the person not only from his or her actions and their source in his or her own convictions, but from his or her self; or is, in other words, to exchange him or her in some way for some other person. One can still wonder what is absurd about this, and Bernard Williams provides no answer to the query.

Most people seem to feel uneasy about the prospect of being exchanged for some other person. The theme is very popular in science fiction and I know of no elaboration of it where the possibility of an exchange is viewed positively. Indeed, it is not difficult to imagine situations where it would not be a good thing to be exchanged for some other person. We may suspect that the person taking our place cannot really feel happiness as we do, or we may suspect that she or he will not be able to fulfil our responsibilities, complete the various tasks we have undertaken and regard as important, and so forth. We suspect that those whom we love and receive love and affection from will not love her or him as they love us. (At the same time we may, out of jealousy, fear that they will!) Suppose, however, that the person about to take our place is precisely like us, with the sole exception that she or he does everything we do a little bit better, is a little bit more happy, and makes the people we love a little bit happier than we do. It is possible that we would nevertheless protest against being exchanged for this person, just as the carpenter in the example protested against the prospect of being exchanged for the better. Could such a protest against being exchanged have any rational foundation? I think not.

If we realise that in fact no one is experiencing any loss (I assume, for the moment, that no transempirical self is lamenting over the loss of a place in time and space), if the exchange as such does not hurt, and if it involves no bad side-effects, then there is a positive obligation on our part to get ourselves exchanged. After all, that is only what takes place when new generations replace old ones. What is so terrible about that?

But whenever a person is changed, as in the example above, or exchanged for another person, there is a lack of felt connectedness, a loss of memory; this, it could be argued, should speak against such changes or exchanges. I think not, however. First of all, the loss of memory may be 'repaired'. The man who has taken the medicine and, consequently, feels no fear may not only be made not to protest against this after the change has taken place; he may also be given the (false) impression (a quasi-memory) that (contrary to fact) he wanted to take the medicine. Secondly, even if there is a lack of mental connectedness after the man has taken the medicine, he need not be worse off in this respect than he was before he took it. Why did he feel such terror of being killed in an accident before he took the medicine? The reason for this may be hidden from him; he

may have forgotten what caused the terror. On the other hand, he now has some understanding of his actual and more relaxed view of the possibility of getting killed. He may come to understand the medical explanation of what happened to him, and so forth. Thirdly, it is not self-evident that it is a condition of a good life to be completely connected by memory. Would it not, on the contrary, be unbearable if we could not forget any single thing that has happened to us? There are some memories which we are only too glad to get rid of. Generally speaking, we cannot live (well) altogether disconnected from memory, and we cannot live (well) completely connected by it either.

However, when one person is exchanged for another, there might be something involved in the transaction which is problematic from an interpersonal point of view. It is often said that when we love a person, we love a unique individual, not a generic type. And the unique individual is lost in the transaction. Does this make the exchange objectionable? I think not. At least the exchange is unobjectionable if we assume, as I have done hitherto, that, in order to be reasonable, the criterion of personal identity must be stated in empirical (physical or mental) terms. Then there is no deep difference between my actual wife and the one I would get if she was exchanged for a different but similar person. When she is exchanged, my feelings need not change; I need not even notice the change. Perhaps my feelings would change, as a matter of fact, if I were told about the exchange. The change in my feelings would be irrational, however. In the situation, nothing has happened that does not happen all the time. The difference between my former and my present wife could be less than the difference between my wife five minutes ago and my wife five years ago. In the former case we may speak of an exchange, because the change is sudden. In the latter case we ought perhaps to speak of change, not of an exchange, since continuity is preserved. However, if the result of the former change (the exchange) is less upsetting than the result of the latter change (which took place over several years), why, then, bother about the way the former change was brought about?

It might be retorted that any validity the foregoing argument may possess only exists because it presupposes much too superficial a view of autonomy – and of personal identity. Admittedly, it might be held that to act autonomously a person must do what he or she decides to do, but this is in fact not sufficient. His or her will must also be autonomous; it must, in other words, be his or her own.

### The autonomous will

How, then, are we to understand the view that the will of an autonomous person must be autonomous?

In the first place, it could be a view such as Kant's that, in order to be autonomous, our actions must spring from neither our empirical selves nor our

actual wants, but our noumenal selves. Only a will with that kind of origin is really our own.

In the second place (setting aside the possibility that there may be, in addition to our empirical selves, noumenal selves), it could be a view to the effect that, in order to be autonomous, our actions must be caused by not our actual wants, but our actual selves; the assumption then is that, even if there are not two kinds of self, yet there are two kinds of causality, namely event causality and agent causality.

In the third place, it could be a view like the one formulated by Harry G. Frankfurt,<sup>14</sup> to the effect that, in order to be autonomous (or 'free', to use Frankfurt's word), our actions must be determined by wants which are in their turn determined by, or at least not at odds with, our second-order volitions, i.e., our volitions about what desires to have, leading to action. This is how Gerald Dworkin conceives of autonomy.<sup>15</sup>

Both the Kantian interpretation, presupposing the existence of two kinds of self, and the interpretation presupposing the existence of two kinds of causality, are problematic from the point of view of ontological economy. Richard Taylor, who used to believe in the doctrine of agent causality, no longer does, and his argument seems to me telling:

what is this 'person' that is supposed to be an originating cause? A person – i.e., a man or woman – is not less a part of the physical world than a clock, or a cat . . . We can therefore suppose that a person is subject to the same kind of change, and exemplifies the same kind of causation, as any other physical object.<sup>16</sup>

Moreover, a special problem with the Kantian interpretation is that, if our noumenal self is causally independent, then it is difficult to see that it can, in any way, be threatened by anything we do to an empirical person. The claim that autonomy must be respected, then, turns out to be vacuous. Isaiah Berlin makes this point in the following way:

Kant's free man needs no public recognition for his inner freedom. If he is treated as a means to some external purpose, that is a wrong act on the part of his exploiters, but his own 'noumenal' status is untouched, and he is fully free, and fully a man, however he may be treated.<sup>17</sup>

A decisive argument against both the Kantian interpretation of 'autonomy' and the interpretation presupposing agent causality is that, even if we may have an obligation to be 'autonomous' in some such rigorous sense (I do not discuss this view in the present context), it is difficult to see that it is of any value to us that we are autonomous; i.e., if we were to lose our 'autonomy', this could mean merely a relief. It must be asked in what way it would be better to have one's actual wishes and decisions controlled by one's noumenal self rather than by

someone else's. In what way would it be better to have one's actual wishes and decisions controlled by any self at all, noumenal or empirical, one's own or someone else's, rather than, say, by one's own previous wishes and decisions and traits of character, with which one is, to some extent at least, connected by memory?

If the problem is, as it is in the present book, what value lies in being autonomous, rather than what kind of obligation there may exist to be autonomous, Frankfurt's (and Dworkin's) concept might seem more promising. However, this concept creates no new complications for my argument. There can be nothing special about second-order volitions. If my first-order volitions are changed, so as to make me happier, it may happen that my second-order volitions 'protest'. But they can be modified to the same extent. This may mean an exchange of persons, if a person's identity is held to be in the integration by his or her second-order personal volitions of his or her first-order volitions. I see, however, no good reason against such an exchange as such. All the reasons adduced against changing someone's second-order volitions seem to be equally good reasons against changing first-order ones. (As a matter of fact, the carpenter's volitions, in the example above, seem to be partly of the second order.) If, then, as I claim to have done, I have disposed of all of the objections against changing these volitions, I have also disposed of the possible objections against changing second-order volitions as well.

### Is it good in itself to be an autonomous person?

Several philosophers have argued that, if we were to give up our autonomy, this would mean that we were giving up not only our own personal identity, but also our mature human nature. Thus J. S. Mill, who argued in 'On Liberty' that 'the free development of individuality is one of the leading essentials of well-being', and who complained that the evil is that 'individual spontaneity is hardly recognised by the common modes of thinking as having any intrinsic worth',<sup>18</sup> states that:

Human nature is not a machine to be built after a model, and set to do exactly the work prescribed for it, but a tree, which requires to grow and develop itself on all sides, according to the tendency of the inward forces which make it a living thing.<sup>19</sup>

Robert Wolff, who argues that men 'are no better than children' if they accept the rule of others, writes as follows:

When I place myself in the hands of another, and permit him to determine the principles by which I shall guide my behaviour, I repudiate the freedom and reason which give me dignity. I am guilty of what Kant might have called the sin of willful heteronomy.<sup>20</sup>

And Joel Feinberg argues along similar lines. He invites us to imagine a person who is given but one course of action in all situations. Such a person, says Feinberg, 'could take no credit or blame' for her or his actions. She or he could have no 'dignity', in her or his own eyes or in the eyes of others. There would be no point in her or his changing her or his mind or purposes. Her or his 'self-monitoring' and 'self-critical' capacities would dry up.<sup>21</sup>

I am not absolutely convinced that the whole of this is true if, after all, this person is always left some course of action. If she or he decides to adopt it, wants to do so and does so, then she or he may well be responsible for her or his actions. But if her or his autonomy is taken away from her or him completely, if all the actions we attribute to her or him are really performed not by herself or himself but by someone else, then she or he is without responsibility, just as Feinberg says that the unfree man is.

How could a person be deprived of his or her autonomy? One possible way of envisaging this outcome could be as follows. The man in question has been the object of brain surgery. The result of this surgery is a complete weakness of this person's will so that thereafter none of his decisions are caused in any way by his will. Instead, his decisions are monitored externally with the aid of a little apparatus implanted in his brain. Sometimes this man decides in accordance with his will, but not because he wills it. In other cases he decides against his will. Afterwards, however, he is made to like his decisions. This man never does what is bad for him except when there are good reasons for doing so, reasons relating to the welfare of others. To avoid irrelevant problems, let us suppose that this person has freely undergone this kind of therapy.

Clearly this man has given up all of his autonomy. He does not perform one autonomous action. He is not that kind of person. He is in the hands of someone controlling him. It is true that he is 'no better than a child' after he has undergone the surgery, and perhaps he is in some respects more like a machine than like a tree.

Is this bad? Is it bad for him to be no better than a child, to lack responsibility, and to be like a machine? Is it bad in itself that he is in this situation? Must we say of the person that if his gain in felt well-being is small enough he has been 'irrational', in the sense that he has chosen a lesser good (or even something evil) before some greater good? I think not.

To be in his situation may be dangerous. He is very vulnerable indeed. However, let us assume, for the sake of argument, that he is well taken care of and that no one takes advantage of his position. Then his choice may be perfectly rational.

To be sure, the patient in my example is without responsibility for his actions; he cannot reasonably be dignified, in his own eyes or in the eyes of others, at least not because of what he has done himself; he cannot change his mind himself; his self-monitoring and self-critical capacities are destroyed. But so what?



Suppose he is happy. Remember that he chose to undergo this therapy. Suppose that, as he expected, he is slightly happier than he was before the operation took place. He might be happy, Feinberg concedes. But a 'contentment with which all this might still be consistent would not be a recognisable human happiness'.<sup>22</sup>

This may be so. Whether this contentment is 'human' or not is a matter of stipulation. However, even if we assume that it is not human, I still cannot see that this creates any serious trouble, for the 'person' experiencing this contentment will not be a human being, either. Feinberg says that he will be like a 'robot' (and Mill used the word 'machine'). Perhaps this is so. But then, why should he complain that his contentment is not human? He most certainly would not. He would not be allowed to complain (we may assume).

However, if it is not bad for him to be in the situation in which he is, how, then, could his being in this situation be bad in itself? It could be maintained, of course, that human contentment is superior to that experienced by non-human sentient beings. 'Man is the final end of creation', wrote Kant. 'Without man the chain of mutually subordinated ends would have no ultimate point of attachment.'<sup>23</sup>

How could such a claim be sustained, however? I know of no good argument in defence of it. But if our preference for human happiness cannot be based on good reasons, it seems to rest on nothing but self-indulgent, gratuitous 'speciesism'. The preference is effectively undermined, it seems to me, by the insight that human nature is the result of a blind process of natural selection; it has slowly become what it is and is slowly changing. The present state of the human gene pool is not sacred.

Nozick's experience machine, described in *Anarchy, State, and Utopia*, now comes to mind. We assume that, if we plug into the machine, then neuro-psychologists stimulate our brains so that we think and feel that we are writing great novels, or making friends, or reading interesting books. All the time we are floating in a tank, with electrodes attached to our brains. If we plug in, or, even more so, if we are manipulated to plug in, we lose all our personal autonomy, I would say. According to Nozick, we do not want to plug into this kind of machine. And there is a lesson to be learnt from this fact: 'We learn that something matters to us in addition to experience by imagining an experience machine and then realizing that we would not use it.'<sup>24</sup> He then goes on to state what it is that matters to us: 'Perhaps what we desire is to live (an active verb) ourselves, in contact with reality.'<sup>25</sup>

The truth of this argument is not unquestionable. Perhaps many people would as a matter of fact opt for the experience machine. And perhaps some people, who would not, would not opt for it because of an (unreasonable) fear that those in charge of the machine would take advantage of them in some nasty way. Moreover, those who hesitate to opt for the machine may do so because they do

not like the fact that the option seems to be irrevocable. This is an unnecessary defect in Nozick's argument, however. He could allow that people now and then become conscious of their present situation; once a year they could be offered the possibility of opting out of the machine. If this is how the machine works, are we to expect that people will not opt for it or that, if they do, they will opt out of it as soon as a possibility of doing so arises?

I am not so sure of this. After all, many people choose to use drugs they know are dangerous, such as alcohol, in spite of the fact that they know that it is difficult to give up the habit of using them. So why not opt for a perfect experience machine (that you can opt out from if you like) with no bad side-effects – and stay plugged into it?

More importantly, however, the argument begs the question. As a matter of fact, by stressing the putative fact that we do not want to plug in, and then using this as an argument to the effect that it would not be a good thing for us if we did, Nozick takes for granted what is highly controversial, and what was rejected in the previous chapter, to wit, that, necessarily, our choices and desires are guides to what is good for us (preferentialism). Nozick seems to be – or to have become – aware of the problem, for later, in *The Examined Life*, he makes the following comment:

Notice that I am not saying simply that since we desire connection to actuality the experience machine is defective because it does not give us whatever we desire . . . . Rather, I am saying that the connection to actuality is important whether or not we desire it – that is *why* we desire it – and the experience machine is inadequate because it doesn't give us *that*.<sup>26</sup>

However, if this is taken seriously, then it emerges that, after all, Nozick does not have any argument for his position. In his famous example he is simply, in a highly dogmatic way, stating it.

Perhaps a similar argument by J. S. Mill fares better, however. According to Mill, there are higher and lower pleasures. Some kinds of pleasure are more desirable and more valuable than others. For example: 'It is better to be a human being dissatisfied than a pig satisfied; better to be Socrates dissatisfied than a fool satisfied.'<sup>27</sup> Perhaps we could argue, in the spirit of Mill, then, that the pleasures felt when our autonomy is gone, when, for example, we are plugged into an experience machine, or live like animals rather than as people, are of less value than the pleasures we feel as autonomous, human beings.

How do we know when we are facing a pleasure of a higher quality, then? 'Of two pleasures, if there be one to which all or almost all who have experience of both give a decided preference, irrespective of any feeling of moral obligation to prefer it, that is the more desirable pleasure.'<sup>28</sup> Once again it seems that preferentialism has been taken for granted. But perhaps we ought to think of

the 'decided preference' not as a mere preference, but as a considered value judgement. If we do then I must say that I sympathise in principle with this test. However, how are we to apply it?

In the first place, I know of no investigation where it has been systematically put to use. It is an open question, then, whether, if it could be made operational, it would yield any result at variance with hedonism. In particular, I conjecture that, if there is no feeling of moral obligation to prefer one pleasure to the other, most people will opt for whatever it is that brings them the most intensive or endurable pleasure, irrespective of whether this is pushpin or poetry, and irrespective of whether they experience it *qua* autonomous individuals or *qua* brains in a vat. They will say that this is the best experience. I conjecture that this is true even of those who prefer being autonomous persons to being brains in a vat.

In the second place, it seems difficult to make the test operational. Suppose we have two persons, having experienced two kinds of pleasure, A and B. Now, the first person claims that A is better, the other one that B is better. What are we to say of the situation? It seems to me that, when faced with a result such as this, we can always deny that these two persons were really having the same kind of experience. Some differences in their perception may account for the difference in their judgement.

The upshot of this is that, even if there is nothing wrong in principle with Mill's argument, it is not decisive one way or the other. If it could be made operational, it might come to support hedonism just as well as the position Mill himself wants to defend.

### Personal autonomy and rule utilitarianism

In the present book I have said little hitherto about rule utilitarianism. But perhaps it could provide a rationale for the belief that we ought to respect people's personal autonomy.

What is rule utilitarianism, then? I have defined it in the following manner:

RU: A particular action is right if, and only if, it is not proscribed by any rule or system of rules such that, if people were generally adhering to it, the world, on the whole, would be better than if they were adhering to any rule or system of rules permitting it.

Now, I have tried to show that personal autonomy lacks intrinsic value. To many this may still seem wrong, or even morally repugnant. I have argued that, as a matter of fact, personal autonomy lacks intrinsic value, but I do not want to argue that it is not morally repugnant to say so. And I admit that the thesis is at variance with some of our moral intuitions. I have tried to show that, if these intuitions are reflected upon, however, if they are put in philosophical

perspective, they have to yield. They do not fit into our 'reflective equilibrium'. This does not mean, however, that these intuitions should be given up. Perhaps they should, but I do not want to argue this point. The fact that they are not epistemologically justified does not mean that they are not morally justified. Our belief that personal autonomy has intrinsic value may have very good consequences in the long run. This means that, on strictly utilitarian grounds, we ought to retain it in our common sense thinking about morality.

I have several times touched upon this idea that our morality must be stratified into a critical and an intuitive level, put forward most famously by R. M. Hare.<sup>29</sup> We must adopt some kind of what Peter Railton has called a 'sophisticated consequentialism'.<sup>30</sup> In Chapter 9 I defend such a stance against strictures directed against it by Jonathan Dancy. If Dancy is right, these two levels cannot cogently be kept separate. If I am right, they can. In the present chapter I will simply take this for granted.

Now, no matter whether Hare is right in his belief that our common sense morality ought to contain a principle of liberty or not, this does not affect my main point that, irrespective of what we ought to believe and to teach our children, personal autonomy does not have intrinsic value.

Against this line of argument it could be retorted, however, that the two levels of moral thinking here discussed should not be kept separate, even if it is possible to do. The argument would run as follows: if a general acceptance of a principle stating that personal autonomy be respected has better consequences than the acceptance of any alternative morality, then this shows not only that such a general acceptance should be sought, but also that, in a particular case, whatever the consequences of doing so, it is right to respect personal autonomy. This is the rule-utilitarian position.

Some commentators on Mill have argued that this is Mill's position. His virtually absolute proscription against interference with personal autonomy is justified, they argue, not because of the intrinsic value of personal autonomy, but rather because of the overall gain in net utility obtained by recognition of such an absolute value.<sup>31</sup>

I do not find this position defensible. I reject RU. It is acceptable to teach a morality, the general acceptance of which has the best consequences, I concede;<sup>32</sup> and, if such teaching has been effective, perhaps, being caught in our common sense morality, we will, as a matter of fact, conform to it, even in situations where such conformity does not produce the best possible outcome; but our conformance in these situations, producing bad effects, is really something we should avoid.

The rule-utilitarian position may seem attractive as long as we do not make a distinction between two levels of moral thinking. However, when we realise that this is something we can do, that it is one thing to question what moral beliefs we ought to have (or to teach to our children), and quite another thing to

question what moral beliefs are true, any rational belief in the rule-utilitarian position gets undermined. To have grasped this distinction fully, and yet to believe that one ought to cling to optimific rules even in situations where by so doing we do not maximise intrinsic value, is to be the victim of what J. J. C. Smart has famously called 'Rule Worship'.<sup>33</sup>

### Conclusion

If the argument of the present chapter is sound, then autonomy lacks intrinsic value. But then all sorts of putative perfectionist value such as achievement, knowledge, deep personal relations, and a contact with realities seem to go overboard as well. For, to be sure, the notion of autonomy is built into these notions.

When it is claimed that achievement, knowledge, deep personal relations, and a contact with realities are of importance as such to the value of the life of an individual, it is taken for granted that these achievements, this knowledge, these deep personal relations, and this contact with realities are truly this person's own. If, from the point of view of intrinsic value, it does not matter whether we make our own achievements, gain our own knowledge, make our own friends, or get ourselves into contact with realities, or are manipulated into these relations, then the relations as such cannot really matter to us.

Perfectionism is not a tenable view.

### Notes

- 1 Or, perhaps we should say that, according to Kant, we cannot help being autonomous. We are, at least as noumenal selves, autonomous by nature.
- 2 J. S. Mill, 'On Liberty', Ch. III.
- 3 J. Rawls, *A Theory of Justice*, p. 516.
- 4 J. Glover, *Causing Death and Saving Lives*, p. 77.
- 5 J. S. Mill, 'On Liberty', pp. 191–2.
- 6 Isaiah Berlin, in 'Two Concepts of Liberty', makes the famous distinction between 'positive' and 'negative' liberty. When I contrast autonomy with liberty/freedom, my concept of autonomy is close to Berlin's concept of positive liberty and my concept of liberty/freedom is close to his concept of negative liberty.
- 7 Rawls insists that the proper word is 'lexicographically', but finds it too cumbersome. He defines a 'lexical order' as an order 'which requires us to satisfy the first principle in the ordering before we can move on to the second, the second before we consider the third, and so on. A principle does not come in to play until those previous to it are either fully met or do not apply' (*A Theory of Justice*, p. 43).
- 8 J. Griffin, *Well-being*, p. 67.
- 9 Griffin, *Well-being*, p. 70.
- 10 P. Singer, *Practical Ethics*, p. 146.

- 11 Singer, *Practical Ethics*, p. 146.
- 12 Glover, *Causing Death and Saving Lives*, p. 82.
- 13 B. Williams, 'A Critique of Utilitarianism', p. 116.
- 14 Cf. H. G. Frankfurt, 'Freedom of the Will and the Concept of a Person'.
- 15 Cf. G. Dworkin, 'The Concept of Autonomy'.
- 16 R. Taylor, 'Agent and Patient', pp. 224–5.
- 17 Berlin, 'Two Concepts of Liberty', p. 156.
- 18 Mill, 'On Liberty', p. 185.
- 19 Mill, 'On Liberty', p. 188.
- 20 R. P. Wolff, *In Defense of Anarchism*, p. 72.
- 21 Cf. J. Feinberg, 'Interest in Liberty on the Scales'.
- 22 Feinberg, 'Interest in Liberty', p. 40.
- 23 I. Kant, *Critique of Judgement*, p. 99.
- 24 R. Nozick, *Anarchy, State, and Utopia*, p. 44.
- 25 Nozick, *Anarchy, State, and Utopia*, p. 45.
- 26 R. Nozick, *The Examined Life: Philosophical Meditations*, pp. 106–7. I owe this observation to Bengt Brülde.
- 27 J. S. Mill, *Utilitarianism*, p. 260.
- 28 Mill, *Utilitarianism*, p. 259.
- 29 Cf. R. M. Hare, *Moral Thinking*, where a theory of two 'levels' of moral thinking is developed, and see Hare, 'What is Wrong With Slavery?', where he defends the claim that our common sense morality ought to contain a 'principle of liberty'.
- 30 Cf. P. Railton, 'Alienation, Consequentialism, and the Demands of Morality', about this.
- 31 Philosophers who have famously discussed this possible interpretation of Mill are J. O. Urmson, in 'The Interpretation of the Philosophy of J. S. Mill', and J. D. Mabbott, in 'Moral Rules'.
- 32 As a matter of fact, the kind of morality we ought to teach, on the most plausible version of the two levels account, is not the one with the best acceptance value, let alone the one with the best conformance value. The one we ought to teach is the one with the best educational value. This is perhaps a morality that people will accept, when it is taught to them, and such that it has better consequences for people to accept it than any other morality that they would accept, if they were taught to do so. Or, it might be a morality such that, if people were taught it, they would reject it, and the consequences of their rejection would be better than their acceptance or rejection of any other morality that they could be taught.
- 33 J. J. C. Smart, 'An Outline of a System of Utilitarian Ethics', p. 10. Notice that, if 'Rule Worship' is indeed a vice, it is an intellectual not a moral one. It may well be the case that, if people practise rule worship, the world is better than if they do not.